The construction of ‘critical thinking’

The Construction of 'Critical Thinking': Between How We Think and What We Believe

Abstract:

'Critical thinking' is widely regarded as important, but difficult to define. This article provides a historical perspective by describing how 'critical thinking' emerged as an object of psychological study, how the forms it took were shaped by practical and social concerns, and how these related to 'critical thinking' as something that results in certain conclusions, rather than as a process of coming to conclusions. 'Critical thinking' became a scientific object when psychologists attempted to measure it. The original measurement treated 'critical thinking' as both an ability and an attitude. It measured logical abilities, and consistency and extremity of views, but it avoided making assumptions about the correctness of specific real-world beliefs. The correctness of such beliefs was, as problems with other related tests showed, open to dispute. Subsequent tests increasingly focused on logical abilities, and attempted to minimize further the relevance of what people believed about the real world, though they continued to depend on there being correct answers to test items, which privileged the outcome over the process. While 'critical thinking' was primarily the domain of philosophers, there was renewed psychological interest in the topic in the 1980s, which increasingly presented 'critical thinking' as incompatible with certain real-world ('unscientific') beliefs. Such a view more explicitly privileged the outcome over the process. It is argued that a more reflective approach, though it may be more difficult to measure, is essential if we wish to understand not only what critical thinking has been, but also what it is now.

‘Critical thinking’ is one of those things that everyone agrees is important. It is widely regarded as a fundamental aim in education, and as a desirable feature of employees and citizens (Sternberg, Roediger, & Halpern, 2007). However, while its importance is generally accepted, there is considerably less agreement about what it is. It has been described, among other things, as a trait, as an attitude, as an ability, and as a combination of more specific skills.[1] Furthermore, while everyone agrees that critical thinking involves thinking, there is no consensus on what kind of thinking is involved. Philosophers have tended to discuss critical thinking as logical thinking, and have focussed on the validity and soundness of arguments, and on the avoidance of logical fallacies. Indeed, for some, ‘critical thinking’ is much the same as Informal Logic (Ennis, 1984; Possin, 2008; Sobocan, 2003). Psychologists, on the other hand, have tended to discuss ‘critical thinking’ as scientific thinking, with greater focus on hypothesis-testing and on an understanding of probability, as part of an attempt to reduce bias and to think in a more objective way (Halpern, 1984; Sternberg, Roediger, and Halpern, 2007). For some, ‘critical thinking’ refers to the proper use of scientific methods (Benjafield, 1994; Meltzoff, 1998). For many psychologists, it as an antidote to ‘erroneous’ (i.e. ‘pseudoscientific’) beliefs (Bensley, 1998; Dunn, Halonen, and Smith, 2008; Halpern, 2014; McBurney, 1996; Ruscio, 2005; Smith, 2010; Stanovich, 2004; Sternberg et al, 2007; Tavris, 2001).

[1] Examples of these are included in this article. The problem of definition is invariably discussed at the beginning of overviews of the topic (for example: Behar-Hornstein and Niu, 2011; Dunn, Halonen, and Smith, 2008; Possin, 2008). More specifically, see: Johnson, 1992; Lipman, 2003, pp. 56-63).

For the most part, however, critical thinking has been concerned with how we think, not with what we believe. For John Dewey, the essence of both 'reflective thinking' and, when he used the term, 'critical thinking', was suspended judgment (Dewey, 1910, pp. 14, 74). It was a process, rather than an outcome. The process included the identification of assumptions and the consideration of alternative positions. Few today would deny that the recognition of assumptions, and the consideration of alternative views, are part of what it means to think critically. Indeed, according to some, this should include identifying the assumptions of, and considering alternatives to, the methods of science (Slife, Reber, and Richardson, 2009; Teo, 2011). In this latter sense, critical thinking, far from seeking to avoid bias in a quest for objectivity, demands the recognition of subjectivity, and scientific thinking becomes the object, rather than the exemplar, of critical thinking.

Given such a diverse range of views, some historical perspective might be useful. The history of critical thinking, of course, could be seen to include the critical thoughts of eminent thinkers since Socrates (Thayer-Bacon, 2000). However, while people may have been thinking in critical ways for millennia, the modern view of 'critical thinking' as something (a trait, an attitude, an ability) that people have, in an amount that can be measured, is a product of the twentieth century. That was when 'critical thinking' became a psychological object, i.e. an object of psychological study (Brock, 2015). This required certain choices to be made about what 'critical thinking' was, particularly in relation to the boundary between how people thought and what they believed.

This article examines the emergence of 'critical thinking' as a psychological object, identifying the assumptions on which it was based, and considering alternatives that were available at the time. It describes how, during the inter-war period in the United States, as psychologists were finding ways to measure what people thought, and ways to change their minds, 'critical thinking' emerged out of growing concerns about how easily the beliefs of

the public could be changed. It shows how the original measurement of critical thinking avoided making assumptions about the correctness of specific real-world beliefs (which was, as problems with other related tests showed, open to dispute), and how subsequent critical thinking tests increasingly focused on logical abilities, though their dependence on there being correct answers to test items privileged the outcome over the process. It also explains how renewed psychological interest in the topic in the 1980s led to critical thinking increasingly being presented as incompatible with certain real-world ('unscientific') beliefs, which more explicitly privileged the outcome over the process. In doing so, it is argued that a more reflective approach is needed to understand not only what critical thinking has been, but also what it is now.

The Psychology of Prejudice and the Concept of 'Fair-mindedness'

In 1925, a psychologist at Columbia, Goodwin Watson, published a new set of tests (Watson, 1925). This was prompted by what he felt was a widespread desire to encourage a certain attitude within students. It was much the same attitude, he explained, as had been described by others as 'open-mindedness', 'scientific-mindedness', and a 'critical, open minded attitude' (p. 2). Watson chose to call it 'fair-mindedness', which he defined as the opposite of prejudice. It was an attitude, he believed, that might be enhanced by appropriate teaching methods. However, to assess which methods were effective, a reliable measure of 'fair-mindedness' was needed. After all, he pointed out, subjective judgments of individual character had been shown to be unreliable, citing the example of the 'halo effect', a general

bias in the rating of individual traits that had been recently identified.[2] Watson's 'Measurement of Fair-mindedness', then, was an attempt to produce an objective measure of individual prejudice, an unbiased measure of the bias of others.

However, while bias in measurement was a problem of accuracy, the bias being measured was of a different kind. Watson was attempting to measure individual prejudice relating to religious and economic issues. His tests of 'fair-mindedness' were not concerned with the correctness of specific views, but rather with openness to the views of others, and with logical consistency in making inferences, judging actions and evaluating arguments in relation to different religious and economic contexts.[3] For example, one might believe in 'faith cures' or not, but to disapprove of them in the context of Roman Catholicism and not in the context of Protestant evangelicalism was treated as a lack of 'fair-mindedness' (p. 10). It is evident that, in Watson's personal opinion, to make gross generalizations (such as 'All Jews would cheat'), or to dismiss as dishonest or incompetent opposing views on debateable issues (such as the success of prohibition), was not only unfair but also wrong. However, even such views were treated as measures of individual prejudice, and as erroneous not in the view that was taken but in the extremity of that position, or in the dismissal of alternative views (p. 11).

---

[2] This was a 'constant error' in interviewers' rating of individuals, which showed a general bias in the ratings of specific individual traits in the direction of the overall rating of the individual (Thorndike, 1920).

[3] There were six tests: Cross-out test (tendency to cross out distasteful words e.g. 'Bolshevik'); Degree of Truth test (degree of certainty on debateable issues e.g. the success of prohibition); Inference test (ability to infer logically from facts); Moral Judgments test (consistency in approval of similar actions in different contexts e.g. faith cures); Arguments test (ability to recognize strong argument, regardless of own view); Generalization test (tendency to generalize e.g. about 'all Jews').

This was not a measure of beliefs, then, but rather of underlying tendencies towards extreme or biased views (whatever these views might be). These tendencies were not seen as general cognitive biases to which the human mind is prone, such as the 'halo effect', but rather as attitudes that individuals might have towards certain groups or ideologies. The battery of tests included those intended to measure abilities that would become a basic feature of subsequent tests of 'critical thinking', such as the ability to infer logically from facts, and to evaluate the strength of arguments. However, Watson's primary concern was how these related to prejudiced attitudes towards religious and economic matters. In short, this was not about false beliefs but about tolerance towards alternative beliefs, and it was not about bias as a cause of inaccuracy but as a lack of open-mindedness.

Watson's interest in 'fair-mindedness' was easy to understand. He had just received his Ph.D. from Teachers College, Columbia, and regarded psychology was a practical way to bring about social change (Nicholson, 1997). The problem of prejudice in the United States, at that time, would have been obvious to him. In the aftermath of the First World War, there was significant anti-German feeling, widespread assumptions about the criminality, radicalism and general inferiority of non-Protestant immigrants, and race riots had occurred in several cities. In 1924, new laws had been passed to restrict immigration from Southern and European countries and, the following year, tens of thousands of Ku Klux Klan activists openly marched in Washington D.C., as representatives of a much broader base. The rhetoric of prejudice was deployed not only by politicians, but also by those psychologists who embraced the hopes and fears of eugenics, who claimed that there was a biological basis for the perceived inferiority of others (Jackson and Weidman, 2004; Richards, 1997). In 1925, a review of studies on 'racial psychology' in *Psychological Bulletin* concluded that they 'seem to indicate the mental superiority of the white race' (Garth, 1925, p. 359).

For those of a more progressive bent, who regarded prejudice as the real problem, measurement was a necessary tool to assess its scale and nature. Indeed, as Watson's tests appeared, a new psychology of prejudice was emerging, along with new forms of measurement (Samelson, 1978). In the same year, Emory Bogardus (1925) published his own measure of prejudice, the Social Distance Scale, and Thurstone's method of paired comparisons was quickly applied to the measure of prejudice (Thurstone, 1928), both of which were attempts to measure prejudice towards racial and ethnic groups. This was part of a growing psychological interest in attitudes of various kinds, and in how they might best be measured, 'attitudes' being defined in a variety of ways, but typically as underlying dispositions that gave rise to surface opinions (Danziger, 1997). Within a few years, hundreds of articles were published on the theory and measurement of attitudes, based on a wide variety of definitions and in relation to a host of social, political and religious views (for contemporary overviews, see: Bain, 1930; Droba, 1932). In particular, social scientists were increasingly attempting to measure and explain prejudice towards certain racial and ethnic groups, 'prejudice' being defined as a negative view of others that was inaccurate and unwarranted (Dixon, 2017).[4]

However, while Watson's 'Measurement of Fair-mindedness' was an early measure of prejudiced attitudes, it did not treat prejudice as necessarily irrational. Watson himself was an advocate of 'fair-mindedness', but he constructed the concept in neutral terms. 'There is no attempt here,' he explained, 'to insist that fair-mindedness rather than prejudice is desirable' (Watson, 1925, p. 7). The neutrality of the construct was linked to its potential utility. The test might be used by those who wished to promote critical open-minded thinking

[4] One can find a similar view in studies by psychologists prior to 1925 (Webster, Saucier, and Harris, 2010)

as an educational objective, but it could also be used in the study of persuasion. Such a measure, Watson claimed, could be useful in assessing 'the effectiveness of lectures, sermons, prayer-meetings, moving-pictures, pageants, group discussions, posters, reading assignments, advertisements, and other supposed techniques for influencing public opinion' (p. 37).

The study of prejudice, as inaccurate and unwarranted, became a significant psychological topic. Watson's concept of 'fair-mindedness' did not. However, by the end of the following decade, the 'fair-mindedness' tests had been revised into something else: the Watson-Glaser Tests of Critical Thinking (1938). This is commonly cited as the first measure of critical thinking and, in its various versions, has long been the most widely used test of its kind (McPeck, 1981, p. 132; Possin, 2014). As we shall see, the transformation of 'fair-mindedness' into 'critical thinking', and the subsequent changes in how the latter was measured, were based on avoiding assumptions about the correctness of specific views. However, it was only one of many attempts by American psychologists in the inter-war period to measure what people thought, and some of these were explicit attempts to address the problem of uncritical, illogical or unscientific thinking. Some were concerned with what people believed; in short, that they believed the wrong things. Assumptions about the correctness of everyday beliefs, however, were problematic.

The Psychology of Belief and the Problem of Correctness

In 1925, as others were in the business of measuring 'attitudes', Harry Hollingworth was attempting to distinguish between 'judgment' and 'belief' (Hollingworth, 1925), a distinction with which others had already struggled (Duncan, 1894; Singer, 1897). Part of the

problem was that, since the beginning of American experimental psychology, the judgments of subjects had been treated in very different ways. On the one hand, experimental subjects had long been required to make judgments about the magnitude of stimuli, but these had been treated as reports of sensory experience, not as studies of accuracy. However, by the end of the century, accuracy of judgment had become a topic of interest, and errors in judgments about the magnitude of measurable phenomena were being attributed to probable errors, personal equations, and other individual differences (Harris, 1915, 1916; Wolfe, 1898a, 1898b; Woodworth and Thorndike, 1900). The potential utility of such work was obvious. If psychologists could identify consistent errors in how people judged distances between objects, for example, then this had 'practical bearings' for everyday practices, from 'the hanging of pictures' to 'the designing of … architectural plans' (Baldwin, 1894, pp. 244-245). After all, as Woodworth and Thorndike (1900) concluded at the end of their experiments, 'in such judgments as we ordinarily make in life there are many factors besides the magnitude of the thing judged which affect the accuracy of the judgment' (p. 355).

However, while the provision of useful psychological knowledge, so essential to the survival of the discipline, depended on coming up with ostensibly objective measurements of socially relevant phenomena, the accuracy of judgments made in the world outside the laboratory was harder to assess. After all, one could attribute errors of judgment of known quantities to specific variables but, as Robert MacDougall (1906) pointed out, these kinds of distortions in judgments were different from the 'sources of unconscious prejudice which influence personal opinion of men and events', since these were affected by unavoidable and unpredictable forms of 'secondary bias', such as selective attention and degree of open-mindedness (p. 100). Furthermore, judgments in ordinary life were not necessarily a question of accuracy. Some psychologists studied judgments of stimuli that were not directly measurable, such as pictures on postcards, handwriting, and jokes, which were treated as

individual preferences (Hollingworth, 1911; Wells, 1908). These were quantified through an 'order of preference' method (for example, jokes were placed into piles according to how funny they were judged to be), and the accuracy of these judgments was not an issue. Thus, when Harry Hollingworth attempted to provide a definition of 'judgment' in 1925, it was obvious to him that it could not be restricted to matters that were either right or wrong, since it included 'verdicts of aesthetics and value' (p. 339).

At the same time, however, there were some beliefs about the everyday world that many simply assumed to be wrong, and which were treated as the product of irrational and unscientific thinking. By then, a psychology of superstition had emerged that bemoaned the persistence of superstitious beliefs among American students, attributing them to a mental predisposition to react emotionally to events, and stressing that education was the only solution (Dresslar, 1907, 1910). To address the problem of superstitious beliefs, teaching methods were sought that reduced superstition, along with ways to measure the efficacy of these methods. In 1925, the introduction of a new method of measurement, which employed a finite list of superstitious beliefs, allowed psychologists to compare the extent of superstitious belief before and after instruction (Nixon, 1925). As the prevalence of such beliefs in schools and colleges appeared to demonstrate the ineffectiveness of general education, many took the view that more specific training was needed (Caldwell and Lundeen, 1932; Dudycha, 1933; Zapf, 1938). Psychology, in particular, was presented as an effective antidote to superstition in studies that claimed that a course in general psychology could, and did, reduce superstitious beliefs (Emme, 1940; Gilliland, 1930; Lehman and Fenton, 1929; Nixon, 1925; Valentine, 1936).

Nevertheless, the assumption that such beliefs were false remained problematic. In one sense, superstitious beliefs could be treated as exemplars of erroneous beliefs, since superstition represented the opposite of scientific knowledge. Indeed, at times, the terms

'superstition', 'unscientific', 'unfounded' and 'false' were used interchangeably. However, in terms of specific beliefs, either their wrongness was simply assumed or else further qualification was provided. For example, H. K. Nixon (1925) noted that some of the beliefs he included 'would occasion considerable discussion among a group of scientific men', but that 'it seems likely that if the list were submitted to a group of psychologists there would be less than 1% of unqualified affirmative answers'; thus, he relied on deviance from a likely consensus of psychologists 'as a rough indicator of the present prevalence of unscientific beliefs' (pp. 419-420). A. R. Gilliland (1930) also admitted that 'a few of the questions may be open to the criticism that they may not really represent superstitions', and that 'many psychologists might mark these questions true' (p. 472). Either way, the beliefs of psychologists were the measuring stick used to judge the correctness (or incorrectness) of such beliefs.

Meanwhile, all of this was in sharp contrast to the wider 'psychology of belief'. From a theoretical perspective, the 'psychology of belief' had been primarily concerned with the relationship of belief to emotion, thought, and other psychological categories (Lindsay, 1910; McDougall, 1921). When experimental studies of belief were carried out, they were concerned with the experience of belief, variations in belief and, particularly, with degrees of conviction (Okabe, 1910; Roback, 1920; Sumner, 1898; Trow, 1923; Williamson, 1915). The stimuli used in these studies were statements, and often included claims about superstitions, religion and psychic phenomena, but responses to these statements were treated merely as a way of accessing degrees of conviction. Indeed, while the role of non-rational factors in shaping beliefs was widely accepted, the experimental 'psychology of belief' was not concerned with whether the beliefs in question were right or wrong. In 1925, Frederick Lund's 'The Psychology of Belief', the first experimental work on the affective determinants of belief, included questions about superstition and survival after death, but the issue of

whether such beliefs were erroneous was simply not raised (Lund, 1925)**.** The practical relevance of the psychology of belief, so far as Lund was concerned, was in explaining how people could be persuaded to believe things (regardless of whether they were true), and those who would benefit from such knowledge, he believed, were 'the advertiser, salesman, journalist, author, politician, minister and lawyer' (p. 63).

In the 1920s, then, psychologists provided a variety of techniques to measure different kinds of thoughts about different sorts of things, and they did so with a market in mind. It was part of a more general attempt to demonstrate the social utility of psychological knowledge, and to supply the growing demand for knowledge about the views of American citizens and consumers (Danziger, 1990, pp. 101-118). On the basis of constructed boundaries between 'judgments', 'beliefs', 'opinions' and 'attitudes', psychologists created seemingly objective measurements of these supposedly discrete phenomena (Danziger, 1997). So far as these measures were of what people thought, of their preferences or of their sense of certainty, these were potentially useful. However, attempts to measure and explain the prevalence of false beliefs about the everyday world reflected an obvious tension between objective methods and social relevance. After all, in the laboratory, one could know the magnitude of the stimulus about which subjects made a judgment, and one could treat deviance from this as error. However, the accuracy of beliefs about the world outside the laboratory was not so easy to measure. One could only assume, as many psychologists did, that certain beliefs were incorrect, and then treat deviance from this view as error. However, even in the case of superstitious beliefs, some clearly recognized this as a problem.

The Psychology of Persuasion and the Problem of Vulnerability

The utility of research on superstitious beliefs was presented in terms of the measurement and reduction of false beliefs, on the basis that such beliefs were not only wrong but also harmful. It was about assessing the extent to which people held these beliefs, and how to change their minds. This was hardly, by any definition then or since, 'critical thinking'. However, although assessing the correctness of beliefs might be problematic, there would always be a demand for knowledge about how to change beliefs. A psychology of persuasion was already well-established, and would continue throughout the inter-war period, but a demand for knowledge about how to resist the techniques of persuasion would also appear. It was out of this latter context that the Watson-Glaser tests of 'critical thinking' emerged.

Since the turn of the century, psychologists had been presenting themselves as experts in persuasion. In the wake of the controversy about the nature of hypnotism, psychologists had taken a keen interest not only in the process of suggestion, but also in the suggestibility of people. There had been various attempts to distinguish between different kinds of suggestion, a concept that was flexible enough to be divided into various slices: there was 'passive' suggestion and 'active' suggestion, 'contrary suggestion' and 'auto-suggestion', and sometimes 'suggestion' was defined 'so broadly that it includes practically all conscious and "unconscious" cognitive and affective processes' (Scott, 1912, p. 270). Meanwhile, experiments had been conducted to discover what kinds of people were most suggestible (Binet, 1900; Sidis, 1898). The concept of 'suggestibility', however, transformed the object of study from an inter-individual process into an intra-individual trait (Danziger, 2008, p. 194). Suggestibility was something that one had, some people had it more than others, and less was preferable to more. It was a kind of vulnerability, which could be described as identical to the 'uncritical self' (Sidis, 1898, p. 364), or as the opposite of 'the critical attitude' (Scott, 1911, p. 310).

Suggestibility was also something that could be measured, and while some thought that a reliable measure might be useful in assessing the reliability of witnesses in the courtroom, there was greater interest in the psychology of persuasion, where suggestibility was something to be exploited (Benjamin, 2007; Kuna, 1976). Walter Dill Scott, a Psychologist at Northwestern University and a former PhD student of Wilhelm Wundt, had stressed the power of suggestion over reasoned argument in advertisements (Scott, 1903, 1908). Harlow Gale, another student of Wundt, had conducted some empirical work on brand associations and attention-getting aspects of advertisements (Eighmey and Sar, 2007), and further experimental work on the efficacy of advertisements had followed (Hollingworth, 1911, 1913; Starch, 1914; Strong, 1911). The Committee on Public Information, set up in 1917 to unite public opinion towards the American war effort, had embraced the psychological techniques of advertisers and public relations experts, such as Freud's nephew Edward Bernays, whose work drew on crowd psychology (Bernays, 1928).

However, in the aftermath of the First World War, concerns were growing about the extent to which public opinion could be influenced. Fears about the possible impact of radical and communist propaganda, which were at their height during the 'First Red Scare', persisted throughout the inter-war period, as the techniques of persuasion could obviously be used for different ends. Raymond Dodge, a psychologist at Wesleyan University, noted that propaganda had become 'a permanent addition to our social and political liabilities', and that it 'was difficult to draw any clear-cut line between advertising and propaganda', though he felt that the latter was more sinister because it 'tends to hide its nature and intention' (Dodge, 1920, pp. 241-2). The desirability of a public that was less suggestible, and more critical, was gradually becoming more obvious.

Throughout the inter-war period, many expressed their concerns about the vulnerability of the American public mind to propaganda (Sproule, 1997, pp. 92-128). Walter

Lippmann, the influential political commentator, described the individual living in the complexity of modern, urban society as someone whose necessarily subjective view of the world could be easily manipulated (Lippmann, 1922). It was a view that concerned Dewey, though he had a more optimistic view of the ability of educated citizens to think for themselves (Dewey, 1922). It was also a matter of concern to psychologists. That year, E. K. Strong's vice-presidential address to the American Association for the Advancement of Science was entitled: 'Control of propaganda as a psychological problem'. He pointed out that advertising and propaganda might differ in their objects and methods, but that both depended on non-rational influences via suggestion, and that by arousing desires and then presenting certain actions as ways to satisfy such desires, people could be persuaded to do things that they would not normally do (Strong, 1922). It was a technique that he would later present as an effective strategy in advertising (Strong and Loveless, 1926). However, in terms of controlling propaganda, he offered no practical solutions.[5]

Psychologists continued to carry out research into what was effective in advertising, though they continued to struggle to find a clear-cut line between advertising and propaganda (Biddle, 1931). However, one clear effect of the growing concerns about the influence of propaganda was a desire among some psychologists to discover how best to resist it. William Biddle of Teachers College, Columbia, whose mentors included Watson, described propaganda as a method of social control that had to be resisted by improving education, specifically by the teaching of critical thinking (Biddle, 1932). By 1935, the use of propaganda techniques in Hitler's rise to power were already the subject of a major study by Leonard Doob at Yale (Doob, 1935). Such concerns, of course, were shared more widely, and

[5] Sceptical about the value of educating the public in propaganda techniques, his only suggestion was "to regulate in some way the use of phrases arousing emotions" (p. 252).

led to the establishment of the Institute for Propaganda Analysis. The institute was set up in 1937 to counter what many saw as the anti-democratic effects of propaganda, and attracted a wide range of progressive academics, including like-minded psychologists such as Doob and Watson (Sproule, 1997). The founder and director of the institute, Clyde Miller, was a progressive educationalist at Teachers College, where Watson worked and where his 'Measurement of Fair-mindedness' had been published. The following year, the Institute of Propaganda Analysis published another set of tests, created by Watson and his student, Edward Glaser.

Measuring 'Critical Thinking'

The Watson-Glaser Tests of Critical Thinking (1938) were a revision of the tests used in Watson's 'Measurement of Fair-mindedness'. They were, like the latter, presented as something that was needed to assess the efficacy of teaching methods (in this case, those which best improved 'critical thinking'). The battery of tests was not presented, as its predecessor had been, as a tool for assessing the effectiveness of techniques for influencing public opinion. On the contrary, as a publication of the Institute for Propaganda Analysis, it was part of a wider attempt to counter techniques of persuasion, so that individuals could think for themselves. As Glaser subsequently pointed out, it was essential that citizens in a democracy could not only read, but also 'think critically' about what they read 'about issues concerning which there may be an honest (or even a dishonest) difference of opinion' (Glaser, 1941, pp. 3-5). The main difference between this set of tests and the measurement of 'fair-mindedness', however, was that two of the tests had been removed, and the remaining ones had been expanded, the result being that it was now primarily an assessment of

reasoning ability.[6] This was in line with Watson and Glaser's definition of 'critical thinking': while 'fair-mindedness' had been defined as an attitude, 'critical thinking' was defined as both an attitude and an ability.

'Critical thinking', as an ability, was measured by testing logical skills. This was not only an understandable decision, but also a practical one. After all, one could test such skills, teach the students how to do logic, and then test them again. So far as 'critical thinking' equated to the ability to demonstrate logical skills, one could expect a significant improvement. Indeed, later that year, the tests were used in schools in New York and New Jersey to measure levels of 'critical thinking' before and after a series of lessons. The series, which was designed to increase 'critical thinking', included several lessons in logic. The result, perhaps not surprisingly, was that students in this experimental group showed significantly greater gains in 'critical thinking ability' than those in the control group (Glaser, 1941, p. 175).

This kind of result, which exemplified the value of the concept, relied on choosing a practical method that reflected certain assumptions. After all, mental testing had been around since the beginning of the century, and intelligence tests were now widely being used in education. However, intelligence was regarded as a fixed attribute of individuals. It was measured by – indeed, it literally was - an individual performance on a test, compared to the performances of others (represented by a statistical norm). Critical thinking, in contrast, was constructed as an ability that could (and should) be improved by appropriate forms of

[6] The two tests that were removed related to tolerance and certainty of attitudes. The test of consistency in attitudes was split into two parts, each consisting of paired opposing statements. The Generalization test, the Inference test (ability to infer logically from facts) and the Arguments test (ability to recognize a strong argument) were complemented by a further test of the ability to infer logically from statements. See appendix for an outline of these tests.

teaching. Evidence of improvement was essential, therefore, if the concept was to survive. The method used in this experiment was a 'classroom experiment', which compared the performances of groups in different teaching conditions, and attributed differences in performance to the form of teaching used. [7] Thus, while critical thinking ability might be something that individuals had, what was really being tested here was neither individual nor group ability, but rather the form of teaching.

So far as Watson and Glaser attempted to measure 'critical thinking' as an attitude, they relied primarily on tests like some of those used in Watson's 'Measurement of Fair-mindedness'. Of the remaining tests that asked subjects about their everyday beliefs, one was concerned with consistency rather than with correctness (it consisted of two parts, each containing one of a pair of opposing statements about, for example, the fairness of wages), and the other test was, like its predecessor, concerned primarily with extremity of views (e.g. 'All Jews would try to cheat a man in a business deal'). In other words, like 'fair-mindedness', beyond gross generalizations about groups of people, 'critical thinking' was not associated with the correctness of specific beliefs about people and events in the real world. Test responses were marked as correct or incorrect, of course, based on the consensus of a panel of judges about what counted as valid conclusions, implicit assumptions, strong or weak arguments, inconsistent beliefs, and over-generalizations, on which the panel was virtually unanimous. Indeed, where disagreement occurred, items were revised until consensus was reached (Glaser, 1941, p. 92).

Such a consensus was essential in a test that claimed to be objective. It was, of course, relatively easy to obtain agreement on how the rules of logic should be applied, and a gross

[7] On the emergence of the classroom experiment and its practical application, see: Danziger, 1990, p. 106ff.).

generalization about a group of people was not particularly hard to spot (since it included words such as ‘all’ or ‘none’). What such responses *represented*, on the other hand, was a more subjective matter. They might be treated, after all, as indicators of ‘prejudice’, of a lack of ‘fair-mindedness’, or of a lack of ‘critical thinking’. Watson’s ‘fair-mindedness’ had itself referred to what others had described, among other things, as a ‘critical, open-minded attitude’ and as ‘scientific mindedness’. The translation of such terms into an objective measure depended on making decisions about what these terms meant, and how, precisely, one could measure the thing to which they referred. If everyone agreed about this, of course, then the measure appeared to be objective.

However, when different choices were made, consensus was harder to find. For example, others had recently tried to provide a measure of the ‘scientific attitude’, which clearly overlapped with these concepts, and which was prompted by a similar desire to instil a more critical attitude in students, but was measured in a quite different way. In doing so, they had struggled to distinguish between ‘scientific thinking’ and ‘scientific attitude’ and, more fundamentally, between ‘attitude’ and ‘aptitude’ (Hoff, 1936; Noll, 1935; Zyve, 1927). The attempt that focused most explicitly on attitude was that of Victor Noll, who defined the ‘scientific attitude’ as a combination of accuracy, intellectual honesty, open-mindedness, suspended judgment, the search for true causal relationships, and criticalness (Noll, 1935). Noting that hundreds of articles had been published on the measurement of attitudes, yet no measure of the ‘scientific attitude’ existed, Noll used items that focused primarily on real-world situations, because he felt that the scientific attitude should be measured in terms of everyday experiences. For example, ‘criticalness’ and ‘open-mindedness’ were measured, respectively, by items such as: ‘If one of my teachers says a thing is so, it must be so’; and ‘Any nation that persecutes the Jews, as Germany has recently done, must be totally

uncivilized' (Noll, 1935, p. 150). In doing so, Noll had to make assumptions about what counted as a correct 'scientific' view about such things, and these were open to dispute.

Indeed, when the test was later administered to several scientists, it was found that 'in numerous instances the scientists do not agree with the published answers', and that in such instances, a child who marked his test in the same way as the scientists 'would be considered unscientific in his thinking' (Blair, 1940, p. 55). On questions related to 'accuracy', which were based on observations and calculations of data provided in the test, 'there was virtually no disagreement' (p. 59). However, on all the other sections, the disagreement was such that the removal of controversial items would have reduced the test to an inadequate length. The only suggestion for improvement was that the test be revised so that 'it would fall in line with the answers of the group of scientists' (p. 59). Such a move, of course, would merely have re-defined the 'scientific attitude' as the possession of similar opinions to those held by this group of scientists, rather than those of Noll.

Meanwhile, had anyone asked Watson and Glaser if the concept of 'critical thinking' was similar to that of 'scientific thinking' or a 'scientific attitude', they would have had to agree. Noll's work on 'scientific thinking' was cited as an example of how aspects of 'critical thinking' could be measured, and was included as essential reading for teachers who participated in their initial experiment (Glaser, 1941, p. 12). However, their approach to measuring 'critical thinking' was radically different from that of Noll. The Watson-Glaser tests, which treated 'critical thinking' as both an attitude and an ability, avoided assumptions about the correctness of beliefs, and focused primarily on logical skills, the correct application of which could be assessed in a relatively direct manner. Noll himself had managed to obtain a consensus on what counted as 'accuracy', because this was based on observations and calculations of data provided in the test. These, like the validity of a conclusion based on premises taken to be true, could be easily checked by anyone familiar

with the relevant methods. However, with that exception, he had relied on (what he took to be) the opinions of scientists as a measure of the correctness of beliefs about the everyday world. Like those who studied superstition, he assumed that (what he took to be) the beliefs of scientists were correct, and treated deviance from this as error (and, among other things, as a lack of 'criticalness').

Watson and Glaser's testing of abilities, on the other hand, was less controversial, and located the concept of 'critical thinking' remarkably close to another contemporary term: 'straight thinking'. This was exemplified by books such as Edwin Clarke's *The Art of Straight Thinking (1929)* and Robert Thouless' *Straight and Crooked Thinking (1932)*, which explained how easily our thinking is influenced by emotion and prejudice, and therefore vulnerable to suggestion and fallacious arguments, and which presented the methods of logic and science as a means of countering internal bias and external propaganda. According to Thouless, whose book was also included as essential reading for those involved in teaching 'critical thinking', 'the ideal of straight thinking must be the application of the scientific habit of thought to all our practical problems' (p. 238). However, his book was concerned with the nature of arguments, and avoided making claims about the correctness of specific beliefs. In this sense, 'critical thinking', as a measurable ability and attitude, was far closer to the ability to think 'straight' than it was to Noll's 'scientific attitude'.

From a practical perspective, this proved to be useful. While the Watson-Glaser tests had emerged from the context of propaganda analysis, the entry of the United States into the Second World War dampened enthusiasm for overt criticisms of propaganda techniques. By 1942, when the activities of the Institute for Propaganda Analysis were suspended, the interests of social scientists had shifted towards a more neutral and less controversial 'straight thinking' approach (Sproule, 1997, p. 99). The general aims of propaganda analysis and straight thinking were not dissimilar, but the former focused primarily on the techniques of

powerful interest groups, and on distortions and biases in the media, while the latter focused on how individuals could employ logical and scientific thinking to separate the wheat from the chaff. The difference between the two was not unlike the difference between suggestion and suggestibility: one referred to an interactional social process, the other to an individual psychological feature. In the latter case, by comparing how well people performed on relevant tasks, measurements could be obtained. Thus, despite significant changes in the wider social context, 'critical thinking' as an individual feature that could be measured, without assuming the correctness of specific everyday beliefs, remained a useful psychological object.[8] Nevertheless, the measurement of 'critical thinking' ability was based on questions that had 'correct' answers, and thus equated 'critical thinking' ability with the selection of these correct answers.

How we think versus what we believe: a brief juxtaposition

Over the following decades, 'critical thinking' was increasingly treated as an ability, one that was independent of any specific beliefs about the real world. By 1952, the Watson-Glaser tests had been revised several times, and the remaining tests that asked subjects about their everyday beliefs had been removed.[9] The 'Watson-Glaser Critical Thinking Appraisal' (WGCTA), as it was now known, focused entirely on reasoning skills (categorized as inference, recognition of assumptions, deduction, interpretation and evaluation of arguments)

[8] Indeed, in a broader sense, the importance of a critical, open mind would become increasingly taken-for-granted in post-war America (Cohen-Cole, 2014).

[9] These were the tests that measured consistency of opinions and generalizations about people (see appendix).

(Watson and Glaser, 1952). Despite this, successful performance continued to depend to some extent on 'commonly accepted knowledge', and on the background beliefs and values of those taking the test, and this, along with several other criticisms, was pointed out (Ennis, 1958). In 1962, the philosopher Robert Ennis provided a more sophisticated and influential analysis of the concept of 'critical thinking', which treated it purely as a set of abilities that could be taught and evaluated (Ennis, 1962).[10] He went on to co-create the 'Cornell Critical Thinking Tests', which reflected this definition. Meanwhile, textbooks and courses on 'critical thinking' were dominated by philosophers, who stressed the role of logical skills, and described various kinds of logical fallacy. By 1980, a growing belief that formal logic was irrelevant to everyday argumentation led to the emergence of Informal Logic, which focused on the structure of everyday arguments, and on the identification of logical fallacies, and 'critical thinking' was increasingly described and taught along these lines (Johnson and Blair, 1980; Lipman, 2003).

Naturally, not all philosophers agreed. For some, the focus on skills was misguided. John Passmore argued that 'critical thinking' should not be regarded as a skill, such as the ability to detect fallacies, and that 'teaching to be critical' was not best achieved by pointing out where students deviated from accepted norms. In his view, critical thinking was more like a character trait, though he also spoke of a 'critical spirit' and of a 'critical attitude' (Passmore, 1967). Others questioned the basis of assessment. John McPeck (1981) pointed out that, while questions designed to test deductive reasoning skills did not require the subject

[10] Ennis here defined critical thinking as 'the correct assessing of statements', which has been taken to mean 'being right', rather than following certain procedures 'correctly' (McPeck, 1981, p. 43). However, Ennis warned of the difficulty in applying his criteria to 'judgments about the world of things, men, and events' (p. 143). In other words, whatever was meant by the phrase, he was not assuming the correctness of certain everyday beliefs.

to know anything more than the information that was provided, questions designed to test inductive reasoning were more problematic, because these were supposedly concerned with matters of truth, rather than of validity. The Watson-Glaser appraisal continued to instruct the subject to rely at times on 'commonly accepted knowledge', and the ambiguity of certain terms meant that subjects with different methodological training, or with different personal beliefs, could justify different answers to the same question. Meanwhile, its main rival, the 'Cornell Critical Thinking Tests', attempted to provide all the relevant information within such questions, but the information could be interpreted in different ways. The general problem, according to McPeck, was that 'rational judgment of a non-deductive sort is not restricted to a unique set of rules, nor to singular solutions' (p. 149). In short, it was inappropriate to reduce 'critical thinking' to the identification of the 'correct' answer.

Critical thinking assessments, of course, depended on there being 'correct' answers. The philosophical focus on fallacies equated 'critical thinking' with 'correct' (i.e. logical) thinking, though this was related to the form of arguments, rather than to specific beliefs about the world. However, the view that certain beliefs were incompatible with critical thinking was to become a significant theme of psychological discourse. Following a 1983 report on the quality of education in the United States by the National Commission on Excellence in Education, which described a worrying lack of critical thinking skills among school students, there was a surge of interest in the topic (Brookfield, 1987). This included a new psychological view of 'critical thinking', which focused less on the rules of logic than on the psychology of reasoning, on problem-solving and decision-making, and on knowledge of hypothesis-testing and probability, as ways to avoid error and bias (Halpern, 1984). By then, there was a high level of psychological interest in 'heuristics and biases', which stressed the extent to which humans were flawed in their thinking (compared to probability theory), and which was producing a growing list of biases that psychologists claimed to have

discovered.[11] It also coincided with the rise of the modern skeptical movement, and a new 'psychology of paranormal belief', which had recently emerged out of a revival of interest in the psychology of superstition, and which explained such beliefs in terms of errors and biases (Lamont, 2013).

Increasingly, 'critical thinking' in Psychology was presented along similar lines. In recent decades, psychologists have frequently contrasted 'critical thinking' with a variety of errors and biases that lead to illogical and unscientific conclusions. The focus on a relatively narrow 'method-centred critical thinking' in psychology has been noted by others (Richardson and Slife, 2011; Yancher, Slife and Warne, 2008). However, this has been accompanied by the claim that false (i.e. 'pseudoscientific') beliefs are exemplars of uncritical thinking, and that the prevalence of such beliefs is evidence of the need for critical thinking (Bensley, 1998; Dunn, Halonen, and Smith, 2008; Halpern, 2014; McBurney, 1996; Ruscio, 2005; Smith, 2010; Stanovich, 2004; Sternberg et al, 2007). For the most part, of course, such textbooks are concerned with how to think logically and scientifically, rather than with the debunking of certain beliefs. Nevertheless, despite the standard view that critical thinking is concerned with how we think, not with what we believe, they make it clear that what (some) people believe is inconsistent with critical thinking. In short, it has become common for psychologists to argue that 'critical thinking' is needed because too many people believe the wrong things. What was explicitly avoided in the quest to measure 'critical thinking' has become a common feature of psychological discourse on the topic.

[11] An early overview of this approach was described in Tversky and Kahneman, 1974). By 1980, this article had been cited 227 times in 127 journals, 80% of which were psychology journals (Lopes, 1991).

Conclusion

Critical thinking became an object of scientific study when psychologists began to measure it. The purpose of measurement was to assess what kinds of teaching increased critical thinking, and the need for increased critical thinking was of the utmost importance: it was essential in a representative democracy, and there was simply not enough of it. In a society that was concerned about the power of propaganda, citizens needed to be able to think for themselves, and come to their own conclusions. Thus, when Watson and Glaser created the original measurement of critical thinking, they defined it as an attitude and ability that was not associated with specific beliefs about the world of people and events. As we have seen, critical thinking survived as a measurable object by increasingly keeping its distance from specific everyday beliefs, and focusing instead on the ability to apply the rules of logic and science correctly.

The form that critical thinking took also reflected a more practical problem of finding a way to manage the gap between scientific objectivity and social utility. 'Critical thinking', as a measurable object, had to be transformed into a quantity. This was done by basing the measure on responses to a list of specific test items, with these responses being treated as instances of critical (or uncritical) thinking. Items that tested the ability to apply certain skills correctly was a practical choice, because this could be assessed in a context in which the correct answer (the one that exemplified critical, rather than uncritical, thinking) was known. This was an artificial context of logical forms and hypothetical scenarios, where the relevant rules were explicit, and all the necessary data was provided. It was a quasi-laboratory setting, in which there was a consensus about the correctness of responses, and assessment appeared to be objective.

Meanwhile, the utility of critical thinking existed in the real world, where people believed all sorts of things for all sorts of reasons. These beliefs were responses to living in the uncontrolled world of people and events. However, attempts to assess critical thinking in terms of the correctness of such beliefs were bound to be inadequate. After all, to attribute incorrect beliefs to a lack of critical thinking, one needs to assume that certain beliefs are incorrect, and that they are due to a lack of critical thinking. Such an attribution is clearly unjustified: one might hold an incorrect belief after critical evaluation of the available evidence, just as one might hold a correct belief without any critical thought whatsoever. One also needs to assume that certain beliefs are incorrect, and even in the case of those beliefs described as superstitions, this was not self-evident.

The psychology of superstition, however, was an example of a different approach to the twin concerns of objectivity and utility. Since the late nineteenth century, psychologists had been making the argument that certain beliefs were wrong and harmful, and that they were the product of errors and biases to which the normal mind is prone, while claiming that by being more like a psychological scientist, such beliefs could be avoided (Lamont, 2010). This 'psychology of error' was a common discursive strategy that had long been used to debunk beliefs in psychic phenomena, and in the interests of disciplinary boundary-work (Lamont, 2013). It was used to justify the new discipline, by contrasting the scientific approach of the New Psychology with ordinary thinking (Scripture, 1898, pp. 1-7). As Pettit (2013) has pointed out, the construction of a 'deceivable self' was a basic concept of scientific psychology. Indeed, it continues to be used in Psychology textbooks, which define scientific Psychology in contrast with common sense and pseudoscientific thinking, the latter being prone to error and bias (e.g. Coon, 2004, p. 10; Martin, Carlson, and Buskist, 2010, p. 4; Raygor, 2005, p. 2). For over a century, the 'psychology of error' has presented

psychological science as an objective route to knowledge, and as a useful antidote to harmful false beliefs.

In recent decades, however, it has become part of psychological discourse on critical thinking, which has been frequently presented as an antidote to 'erroneous' beliefs. In this, there is an obvious overlap with mainstream psychological approaches to paranormal belief, which emerged from the psychology of superstition, and with the literature on 'heuristics and biases', which attempts to explain why people come to irrational conclusions.[12] This emphasis on a common theme, that the normal mind is prone to errors and biases that lead us to get things wrong, has been a deliberate strategy, if not an original one (Lamont, 2010, 2013; Lopes, 1991; Pettit, 2013). However, the value of such an approach, in terms of critical thinking, is questionable. After all, if we accept that we are all prone to error and bias, then how do we know which beliefs are correct, and which are not? The answer, we are told, is that critical thinkers base their beliefs on logic and science.

This, of course, is simply inadequate. Philosophers often disagree, and scientists often disagree, because people can be logical and scientific, yet hold quite different beliefs. To a reflective critical thinker, this is neither surprising nor problematic. The problem only arises when one treats certain beliefs as the product of illogical, unscientific or uncritical thinking. Even in the case of the usual suspects, such as claims related to astrology, paranormal phenomena or alternative medicine, there have always been some people, including some scientists, who have argued that such beliefs have empirical support. To describe these beliefs as false, irrational, or pseudoscientific (or as the product of error and bias) depends on making certain assumptions about the nature of truth, reason or science (or about accuracy

[12] The claims of the 'heuristics and biases' approach has also been criticized, among other things, for assuming the problems they pose have only one 'correct' answer (Gigerenzer, 1991, p.86).

and objectivity). Some might regard this as precisely the sort of thing that a critical thinker would wish to consider. Rather than treat certain beliefs as the products of uncritical thinking, one might wish to think more critically about the underlying assumptions on which knowledge claims depend.

More generally, the critical thinking literature continues to reflect ongoing tensions relating to the objective measurement of an ambiguous concept, the utility of which is inseparable from the world of people and events. A variety of assumptions have had to be made about what critical thinking should be, what it is for, and what can be done about it. The attempt to measure critical thinking is based on the assumptions that it can and should be improved, and that, to facilitate this, it can and should be quantified. It is therefore measured in terms of outcomes (i.e. responses to test items) and, perhaps not surprisingly, it has been frequently defined in terms of outcomes (such as finding solutions and making decisions about what to believe and what to do). This is understandable, but it privileges the outcome over the process, and some would say that the process of critical thinking, regardless of the outcome, is at least as important.

The practical utility of critical thinking, of course, is that it is applicable in the real world. We can teach logical and scientific skills, and we can measure improvement in these, but the utility of the concept of 'critical thinking' is based on the view that it amounts to more than domain-specific skills. It is essential, therefore, that any improvement is not restricted to the context in which critical thinking is taught and measured. Ongoing debates about the generalizability of critical thinking continue to reflect such concerns. Meanwhile, a revival of the view that critical thinking is an attitude, as well as an ability, has led to attempts to measure separately its dispositional aspects (Facione and Facione, 1993). The measurement of attitudes or dispositions, however, is based on another set of assumptions, many of which

have long been questioned.[13] Such are the problems of measuring 'critical thinking', and the practical need to do so is unlikely to go away. It depends, of course, on making certain assumptions and decisions. However, a more reflective approach, which questions such assumptions and considers alternatives, can provide a useful perspective not only on how we have thought about critical thinking, but also on how we might think critically now, and in the future. After all, for Dewey, reflective thinking was a process, not an outcome. The quest for measurement implicitly privileged outcome over process and, more recently, psychologists have explicitly privileged certain conclusions about the world over the process of coming to these conclusions. A more reflective approach to critical thinking, while it may be more difficult to measure, is needed if we wish to understand how we come to the conclusions that we do.

References

Bain, R. (1930). Theory and Measurement of Attitudes and Opinions. *Psychological Bulletin*, 27(5), 357-379.

Baldwin, J. M. (1894). The Effect of Size-Contrast Upon Judgments of Position in the Retinal Field. *Psychological Review*, 1(3), 244-245.

Behar-Hornstein, L. S. & Niu, L. (2011). Teaching Critical Thinking Skills in Higher Education: a Review of the Literature. *Journal of College Teaching and Learning*, 8(2), 25-

[13] Skepticism about such assumptions led some to abandon attempts to measure attitudes entirely (Potter and Wetherell, 1987).

41.

Benjafield, J. G. (1994). *Thinking Critically About Research Methods*. Boston: Allyn & Bacon.

Benjamin, L. (2007). *A Brief History of Modern Psychology*. Malden, MA: Blackwell Publishing.

Bensley, D. A. (1998). *Critical thinking in Psychology: a Unified Schools Approach*. Pacific Grove, CA: Thomson Brooks/Cole Publishing Co.

Bernays, E. (1928). *Propaganda*. New York: Horace Liveright.

Biddle, W. (1931). A Psychological Definition of Propaganda. *Journal of Abnormal and Social Psychology*, 26(3), 283-295.

Biddle, W. (1932). *Propaganda and Education.* New York: Teachers College, Columbia.

Binet, A. (1900). *La Suggestibilité.* Paris: Schleicher.

Blair, G. M. (1940). The validity of the Noll test of scientific thinking. *Journal of Educational Psychology*, 31(1), 53-59.

Bogardus, E. (1925). Measuring Social Distance. *Journal of Applied Sociology*, 9(2), 299-308.

Brock, A. (2015). The history of psychological objects. In J. Martin, J. Sugarman, & K Slaney (Ed.), *The Wiley Handbook of Theoretical and Philosophical Psychology: Methods, Approaches, and new Directions for Social Sciences*. Malden, MA: John Wiley & Sons.

Brookfield, S. D. (1987). *Developing Critical Thinkers.* Milton Keynes: Open University press

Caldwell, O. W., & Lundeen, G. E. (1932). *An Experimental Study of Superstitions and Other Unfounded Beliefs*. New York: Teachers College, Columbia.

Cohen-Cole, J. (2014). *The open mind: Cold war politics and the sciences of human nature*. Chicago: University of Chicago Press.

Coon, D. (2004). *Introduction to Psychology: Gateways to Mind and Behavior. 10th edition*. Belmont, CA: Thomas-Wadsworth.

Danziger, K. (1990). *Constructing the Subject: Historical Origins of Psychological Research*. Cambridge: Cambridge University Press.

Danziger, K. (1997). *Naming the Mind*. London: Sage Publications Ltd.

Danziger, K. (2008). *Marking the Mind: A History of Memory*. Cambridge, UK: Cambridge University Press.

Dewey, J. (1910). *How We Think*. Boston: D. C. Heath.

Dewey, J. (1922). Review of Public Opinion. *The New Republic*, 30(387), 286-288.

Dixon, J. (2017). 'Thinking Ill of Others Without Sufficient Warrant?' Transcending the Accuracy-Inaccuracy Dualism in Prejudice and Stereotyping Research. *British Journal of Social Psychology*, 56(1), 4-27.

Dodge, R. (1920). The Psychology of Propaganda. *Religious Education*, 15, 241-53.

Doob, L. W. (1935). *Propaganda: its Psychology and Technique*. New York: Henry Holt and Company, 1935).

Dresslar, F. B. (1907). Superstition and Education. University of Berkeley Press.

Dresslar, F. B. (1910). Suggestions on the Psychology of Superstition. *American Journal of Insanity*, 67(2), 213-226.

Droba, D. C. (1932). Methods for Measuring Attitudes. *Psychological Bulletin*, 29(5), 309-323.

Dudycha, G. J. (1933). The Superstitious Beliefs of College Students. *Journal of Abnormal and Social Psychology*, 27(4), 457-464.

Duncan, G. M. (1894). Judgment and Belief. *Psychological Review*, 1(2), 202-209.

Dunn, D.S., Halonen, J. .S, & Smith, R. A. (2008). Engaging Minds: Introducing Best Practices in Teaching Critical Thinking in Psychology. In D. S. Dunn, J. S. Halonen, & R. A.

Smith (Eds)., *Teaching Critical Thinking in Psychology: a Handbook of Best Practices* (pp. 1-8). Oxford: Blackwell Publishing Ltd.

Eighmey, J. & Sar, S. (2007). Harlow Gale and the Origins of the Psychology of Advertising. *Journal of Advertising*, 36(4), 147-158.

Emme, E. E. (1940). Modification and Origin of Certain Beliefs in Superstition Among 96 College Students. *Journal of Psychology*, 10, 279-291.

Ennis, R. H. (1958). An Appraisal of the Watson-Glaser Critical Thinking Appraisal. *Journal of Educational Research*, 52(4), 155-158.

Ennis, R. H. (1962). A Concept of Critical Thinking: A Proposed Basis for Research in the Teaching and Evaluation of Critical Thinking Ability. *Harvard Educational Review*, 32(81), 81-111.

Ennis, R. (1984). Problems in Testing IL/CT/Reasoning Ability. *Informal Logic*, 6(1), 3–9.

Facione, P. A. & Facione, N. C. (1993). *The California Critical Thinking Dispositions Inventory (CCTDI); and the CCDTA Test Manual*. Millbrae, CA: California Academic Press.

Garth, T. (1925). A Review of Racial Psychology. *Psychological Bulletin*, 22, 343-364.

Gigerenzer, G. (1991). How to Make Cognitive Illusions Disappear: Beyond "Heuristics and Biases. *European Review of Social Psychology*, 2(1), 83-115.

Gilliland, A. R. (1930). A Study of the Superstition of College Students. *Journal of Abnormal and Social Psychology*, 24(4), 472-479.

Glaser, E. (1941). *An Experiment in the Development of Critical Thinking*. New York: Teachers College, Columbia University.

Halpern, D. (1984). *Thought and Knowledge: an Introduction to Critical Thinking*. Hillsdale, NJ: Lawrence Erlbaum.

Halpern, D. (2014). *Thought and Knowledge: an Introduction to Critical Thinking. 5th ed.* New York: Psychology Press.

Harris, J. A. (1915). Experimental Data on Errors of Judgment in the Estimation of the Number of Objects in Moderately Large Samples, with Special Reference to Personal Equation. *Psychological Review* 22(6), 490-511.

Harris, J. A. (1916). Influence of Previous Experience on Personal Equation and Steadiness of Judgment in the Estimation of the Number of Objects in Moderately Large Samples. *Psychological Review*, 23(1), 30-48.

Hoff, A. G. (1936). A Test for Scientific Attitudes. *School Science and Mathematics*, 36, 763-770.

Hollingworth, H. L. (1911). Experimental Studies in Judgment: Judgments of the Comic. *Psychological Review*, 18(2), 132-156.

Hollingworth, H. L. (1913). *Advertising and Selling*. New York: Appleton.

Hollingworth, H. L. (1925). The Definition of Judgment. *Psychological Review*, 32(5), 337-361.

Jackson, J. P. & Weidman, N. (2004). *Race, Racism and Science: Social Impact and Interaction*. Santa Barbara, CA: ABC-CLIO, Inc.

Johnson, R. H. (1992). The Problem of Defining Critical Thinking. In S. Norris (Ed.), *The Generalizability of Critical Thinking* (pp. 38-53). New York: Teachers College Press.

Johnson, R. H. & Blair, A. J. (1980). The Recent Development of Informal Logic. *Informal Logic: The First International Symposium*. San Francisco: Edgepress.

Kuna, D. P. (1976). The Concept of Suggestion in the Early History of Advertising Psychology. *Journal of the History of the Behavioral Sciences*, 12(4), 347-353.

Lamont, P. (2010). Debunking and the Psychology of Error: a Historical Analysis of Psychological Matters. *Qualitative Research in Psychology*, 7(1), 34-44.

Lamont, P. (2013). *Extraordinary Beliefs: A Historical Approach to a Psychological Problem*. Cambridge: Cambridge University Press.

Lehman H. C. & Fenton, N. (1929). The Prevalence of Certain Misconceptions and Superstitions Among College Students Before and After a Course in Psychology. *Education*, 50, 485-494.

Lindsay, J. (1910). *The Psychology of Belief*. Edinburgh: William Blackwood and Son.

Lipman, M. S. (2003). *Thinking in education.* Cambridge: Cambridge University Press.

Lippmann, W. (1922). *Public Opinion.* New York: Harcourt, Brace and Company.

Lopes, L. (1991). The Rhetoric of Irrationality. *Theory and Psychology*, 1(1), 65-82.

Lund, F. L. (1925). The Psychology of Belief: a Study of its Emotional and Volitional Determinants. *Journal of Abnormal and Social Psychology*, 20(1), 63-81, 20(2), 174-96.

MacDougall, R. (1906). On secondary bias in objective judgments. *Psychological Review*, 13(2), 97-120.

Martin, G. N., Carlson, N. R. & Buskist, W. (2010). *Psychology. 4th edition*. Harlow: Pearson Education Limited.

McBurney, D. H. (1996). *How To Think Like a Psychologist: Critical Thinking in Psychology.* Upper Saddle River, NJ: Prentice-Hall Inc.

McDougall, W. (1921). Belief as a Derived Emotion. *Psychological Review*, 28(5), 315-328.

McPeck, J. (1981). *Critical Thinking and Education*. Oxford: Martin Roberson and Company Ltd.

Meltzoff, J. (1998). *Critical Thinking About Research: Psychology and Related Fields*. Washington, DC: American Psychological Association.

Nicholson, I. (1997). The Politics of Scientific Social Reform, 1936-1960: Goodwin Watson and the Society for the Psychological Study of Social Issues. *Journal of the History of the Behavioral Sciences*, 33(1), 39-60.

Nixon, H. K. (1925). Popular answers to some psychological questions. *American Journal of Psychology*, 36(3), 418-23.

Noll, V. H. (1935). Measuring the scientific attitude. *Journal of Abnormal and Social Psychology*, 30(2), 145-54.

Okabe, T. (1910). An experimental study of belief. *American Journal of Psychology*, 21(4), 563-596.

Passmore, J. (1967). On teaching to be critical. In R. S. Peters (Ed.), *The Concept of Education* (pp. 192-211). London: Routledge and Kegan Paul.

Pettit, M. (2013). *The Science of Deception: Psychology and Commerce in America*. Chicago: University of Chicago Press.

Possin, K. (2008). A Field Guide to Critical-Thinking Assessment. *Teaching Philosophy*, 31(3), 201-228.

Possin, K. (2014). Critique of the Watson-Glaser Critical Thinking Appraisal Test: the More You Know, the Lower Your Score. *Informal Logic*, 34(4), 393-416.

Potter J. & Wetherell, M. (1987). *Discourse and Social Psychology: Beyond Attitudes and Behaviour*. London: Sage Publications.

Raygor, R. (2005). *The Science of Psychology*. Harcourt College Publishers.

Richards, G. (1997). *Race, Racism and Psychology: Towards a Reflexive History*. London: Routledge.

Richardson, F. C. & Slife, B. D. (2011). Critical thinking in social and psychological inquiry. *Journal of Theoretical and Philosophical Psychology*, 31(3), 165-172.

Roback, A. (1920). The Psychology of Belief. *Psychological Bulletin*, 17(2), 53-54.

Ruscio, J. (2005). *Critical Thinking in Psychology*: *Separating Sense from Nonsense*. Belmont, CA: Wadsworth Publishing.

Samelson, F. (1978). From "race psychology" to "studies in prejudice": Some observations on the thematic reversal in social psychology. *Journal of the History of the Behavioural Sciences,* 14, 265-278.

Scott, W. D. (1912). Suggestion. *Psychological Bulletin*, 9(7), 269-271.

Scripture, E. W. (1898). *The New Psychology*. London: Walter Scott Ltd.

Sidis, B. (1898). *The Psychology of Suggestion*. New York: Appleton and Co.

Singer, E. A. (1897). Studies in Sensation and Judgment. *Psychological Review*, 4(3), 250-271.

Slife, B. D., Reber, J. S., & Richardson, F. C. (2009). *Critical Thinking about Psychology: Hidden Assumptions and Plausible Alternatives*. Washington, DC: American Psychological Association.

Smith, J. C. (2010). *Pseudoscience and Extraordinary Claims of the Paranormal: a Critical Thinker's Toolkit*. Malden, MA: Wiley-Blackwell.

Sobocan, J. (2003). Teaching Informal Logic and Critical Thinking. Available at http://web2.uwindsor.ca/faculty/arts/philosophy/ILat25/ papers.htm.

Sproule, J. M. (1997). *Propaganda and Democracy: the American Experience of Media and Mass Persuasion.* Cambridge: Cambridge University Press.

Stanovich, K. E. (2004). *How To Think Straight About Psychology*. Boston: Allyn and Bacon.

Starch, D. (1914). *Advertising: its Principles, Practice, and Technique*. Chicago: Scott, Foresman.

Sternberg, R. J., Roediger, H. L., & Halpern, D. (2007). *Critical Thinking in Psychology*. Cambridge: Cambridge University Press.

Strong, E. K. (1922). Control of propaganda as a psychological problem. *The Scientific Monthly*, 14(3), 234-252.

Strong, E. K. & Loveless, J. E. (1926). "Want" and "Solution" Advertisements. *Journal of Applied Psychology*, 10(3), 346-366.

Sumner, F. B. (1898). A Statistical Study of Belief. *Psychological Review*, 5(6), 616-31.

Tavris, C. (2001). *Psychobabble and biobunk: Using Psychological Science to Think*

Critically About Issues in the News. Upper Saddle River, NJ: Prentice-Hall.

Teo, T. (2011). Radical philosophical critique and critical thinking in psychology. *Journal of Theoretical and Philosophical Psychology*, 31(3), 193-199.

Thayer-Bacon, B. J. (2000). *Transforming critical thinking: thinking constructively*. New York: Teachers College Press.

Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology*, 4(1), 25-29.

Thouless, R. H. (1932). *Straight and Crooked Thinking*. New York: Simon and Schuster.

Thurstone, L. L. (1928). An Experimental Study of Nationality Preferences. *Journal of General Psychology*, 1(3), 405-425.

Trow, W. C. (1923). *The Psychology of Confidence: an Experimental Enquiry. Archives of Psychology, no. 67*. London: G. E. Stechert and Co.

Tversky, A. & Kahneman, D. (1974). Judgment Under Uncertainty: Heuristics and Biases. *Science*, 185, 1124-1131.

Valentine, W. L. (1936). Common Misconceptions of College Students. *Journal of Applied Psychology*, 20(6), 633-658.

Watson G. & Glaser, E. (1952). *Watson-Glaser Critical Thinking Appraisal*. New York: World Book Co.

Watson, G. (1925). *The Measurement of Fair-mindedness*. New York: Teachers College, Columbia University.

Webster, R. J., Saucier, D. A. , & Harris, R. J. (2010). Before the Measurement of Prejudice: Early Psychological and Sociological Papers on Prejudice. *Journal of the History of the Behavioural Sciences*, 46(3), 300-313.

Wells, F. L. (1908). On the Variability of Individual Judgment. In E. L. Thorndike (Ed.), *Essays Philosophical and Psychological in Honor of William James by his Colleagues at Columbia University* (pp. 511-549). New York: Longmans, Green.

Williamson, G. F. (1915). Individual Differences in Belief, Measured and Expressed by Degrees of Confidence. *Journal of Philosophy, Psychology and Scientific Methods*, 12, 127-137.

Wolfe, H. K. (1898a). Some Judgments on the Size of Familiar Objects. *American Journal of Psychology*, 9(2), 137.

Wolfe, H. K. (1898b). Some Effects of Size on Judgments of Weight. *Psychological Review*, 5(1), 25-54.

Woodworth, R.S. & Thorndike, e. (1900). Judgments of Magnitude by Comparison with a Mental Standard. *Psychological Review,* 7(4), 344-355.

Yanchar, S. C., Slife, B. D., & Warne, R. (2008). Critical thinking as disciplinary practice. *Review of General Psychology*, 12(3), 265-281.

Zapf, R. M. (1938). Superstitions of Junior High School Pupils part II: Effect of Instruction on Superstitious Beliefs. *Journal of Educational Research*, 31(7), 481-496.

Zyve, D. L. (1927). A Test of Scientific Aptitudes. *Journal of Educational Psychology*, 18(8), 525-546.

APPENDIX: The Structure of the Watson and Glaser tests

Watson's Test of Fair-mindedness (1925)

This consisted of six tests:

Cross-out test [this measured the tendency to cross out annoying or distasteful words, such as 'Bolshevik']

Degree of Truth test [this measured the degree of certainty on debateable issues, such as the failure of Prohibition];

Inference test [this measured the ability to infer logically from facts]

Moral Judgments test [this measured the consistency in approval/disapproval of similar actions in different contexts, such as faith cures in the context of Roman Catholicism and Protestant evangelism]

Arguments test [this measured the ability to distinguish between strong and weak arguments, regardless of respondent's own position about, for example, the Ku Klux Klan]

Generalization test [this measured the ability to generalize appropriately, particularly not to over-generalize about, for example, 'All Jews']

The Watson-Glaser Tests of Critical Thinking (1938)

These were described as an 'extensive revision' (Glaser, 1941, 87) of Watson's Test of Fair-mindedness (1925), and also consisted of six tests: the Cross-out test and Degree of Truth test were removed; the other tests were revised and expanded, with greater emphasis on logical thinking:

Test A: Survey of Opinions [this was in two parts, and measured consistency of opinions by comparing responses to the first and second part, each consisting of paired opposing statements about, for example the fairness of wages in society]

Test B: General Logical Reasoning [this measured the ability to infer logically from facts, and consisted of syllogistic exercises and hypothetical scenarios]

Test C: Inference Test [this measured the ability to judge the probable truth or falsity and relevancy of inferences drawn from given statements of fact]

Test D: Generalization Test [this measured the ability to generalize appropriately, particularly not to over-generalize about, for example, 'All Jews'].

Test E: The Discrimination of Arguments [this measured the ability to distinguish between strong and weak arguments, regardless of respondent's own position].

Test F: Evaluation of arguments [this measured the ability to identify a logical conclusion from information provided in a statement].

The 'Watson-Glaser Critical Thinking Appraisal' (1952)

This focused entirely on reasoning skills, divided into the following (somewhat ambiguous) categories :

Test 1: Inference [this measured the ability to judge the probable truth or falsity of inferences drawn from given statements of fact]

Test 2: Recognition of assumptions [this measured the ability to identify assumptions implicit in given statements]

Test 3: Deduction [this measured the ability to infer logically from given statements]

Test 4: Interpretation [this measured the ability to identify valid conclusions from premises]

Test 5: Evaluation of arguments [this measured the ability to distinguish between strong and weak arguments]